

TARTU ÜLIKOOL
MATEMAATIKA-INFORMAATIKATEADUSKOND
Arvutiteaduse instituut
Tarkvaratehnika õppekava

Siim Halapuu

Eluslooduse klassifikatsioonide haldamise andmebaas ja veebiliides

Magistritöö (30 EAP)

Juhendaja: Kessy Abarenkov, PhD

Juhendaja: Vambola Leping, MSc

Eluslooduse klassifikatsioonide haldamise andmebaas ja veebiliides

Lühikokkuvõte:

Käesolevas töös käsitletakse eluslooduse klassifikatsioone haldavaid tarkvarasid ning arendatakse veebiteenus bioloogiliste taksonoomiate paremaks haldamiseks. Töö annab ülevaate eluslooduse klassifikatsioonide arendamisega seotud terminoloogiast ja protsessist ning võrdleb olemasolevaid haldusliideseid.

Töös kirjeldatakse erinevaid võimalusi hierarhiliste andmete talletamiseks relatsioonilises andmebaasis ning valitakse sobivaim viis muutliku eluslooduse klassifikatsiooni talletamiseks. Valitud andmebaasi disainimustrit kasutatakse REST stiilis veebiliidese arendamisel. Luuakse kõiki kasutajapoolseid nõuded täitev veebiteenus kasutades Django REST raamistikku.

Magistritöö ühe praktilise osana testitakse loodud veebiliidese enimkasutatud sihtpunktide kasutuskiirust erineva hulga algandmetega. Lisaks antakse soovitusi rakenduse kasutuskiiiruse ja -mugavuse edasiseks parandamiseks.

Võtmesõnad:

veebiteenus, REST API, taksonoomia, klassifikatsioon, elurikkus, hierarhilised andmed relatsioonilises andmebaasis, sulundi tabel, PlutoF

Database and web interface for managing multiple biological classifications

Abstract:

This master's thesis reviews software that is built for managing biological classifications and builds a new web service for better management of multiple biological classifications. The thesis gives an overview about terminology and the process of developing biological classifications.

The thesis describes different possibilities for storing hierarchical data in a relational database. The most suitable method is used for storing constantly changing biological classifications. The chosen database design pattern is used when building the REST web service with Django REST framework. The web service meets all client specified requirements.

As part of the practical work, performance test are ran against most used API endpoints. The performance tests are ran with different sized classifications. Additionally the author gives recommendations for further improving user experience and module performance.

Keywords:

web service, REST API, taxonomy, classification, biodiversity, hierarchical data in relational database, closure table, PlutoF

Sisukord

Sissejuhatus	6
1 Taust	8
1.1 Taksonoomia töö meetodid	11
1.2 Taksonite nimetamine	12
1.3 Taksonoomilised andmebaasid	13
1.4 Klassifikatsioonide haldamine ja kasutamine	13
1.5 Motivatsioon	14
1.6 Kasutajad	15
1.7 Kasutajapoolsed nõuded	16
2 Ülevaade rakendustest	17
2.1 Global Names Architecture	18
2.2 EDIT Platform for CyberTaxonomy	19
2.3 Scratchpads biodiversity online	20
2.4 Tulemused	21
3 Väljapakutud lahendus	23
3.1 Kasutatud tarkvaraarenduse võtted	23
3.2 Andmemudel	24
3.3 Veebiteenus	27
3.3.1 Aktide filtreerimine	30
3.3.2 Alamklassifikatsiooni pärimine	30
3.3.3 Kõigi vanemate pärimine	30
3.3.4 Kattuv klassifikatsiooni tee	31
3.3.5 Otseste järglaste pärimine	31

3.3.6	Taksonite filtreerimine	31
3.3.7	Taksonite otsimine	32
3.4	Automatiseeritud taustategevused	32
3.5	Kasutuskiiirus	33
3.6	Rakendusvõimalused	34
4	Edasine töö	37
	Kokkuvõte	38
	Kasutatud kirjandus	40
	Lisad	43
I.	Terminid	43
II.	Veebiteenuse sihtpunktid	45
III.	Veebiteenuse sihtpunktide väljade kirjeldused	48
IV.	Kasutuskiiiruse tulemused	51
V.	Litsents	52

Sissejuhatus

Bioloogiliste uurimuste üks kommunikatsiooni alustalasid on eluslooduse klassifikatsioon. Selle järgi määratakse, millised liigid kuuluvad samasse perekonda, sugukonda, seltsi jne. Maailmas leidub sadu, kui mitte tuhandeid erinevaid eluslooduse klassifikatsioone [1], mis erinevad üksteisest mitmete tunnuste, nagu näiteks liigi ja kõrgemal tasemel taksonite arvu, liikide grupeerimise kui ka taksonite tasemete poolest.

Taksonoomid on tihti eriarvamusel, milline on parim klassifikatsioon teatud organismirühmade jaoks [2]. Sõltuvalt kasutatud tunnustest ja arvutusmeetoditest saadakse uurimistöö tulemusena erinevaid klassifikatsioone, mida saab vaadelda teaduslike hüpoteesidena. Paljud asutused ja andmekogude haldajad talletavad ja arendavad oma eluslooduse klassifikatsiooni. Samas on aga üha enam hakatud mõistma, et on tarvis luua ühtne kogum erinevatest eluslooduse klassifikatsioonidest [2].

Ülaltoodud probleemi lahendamiseks on loodud mitmeid rakendusi, millest olulisemad on EDIT Platform for CyberTaxonomy¹ (EDIT), Scratchpads biodiversity online² (Scratchpads) ja Global Names Architecture³ (GNA). Autorile teadaolevalt antud rakenduste eluslooduse klassifikatsiooni moodulite varasem võrdlev analüüs puudub.

Antud keskkondade eluslooduse klassifikatsiooni moodulitega tutvudes leidis autor puudusi. Käesoleva töö eesmärgiks on luua REST (*Representational State Transfer*)⁴ stiilis veebiliides (veebiteenus) eluslooduse klassifikatsioonide haldamiseks, milles pakutakse lahendusi leitud puudustele.

Käesoleva magistritöö eesmärkideks on:

- kirjeldada Tartu Ülikooli loodusmuuseumile vajamineva taksonoomiamooduli kasutajapoolsed nõuded;
- hinnata olemasolevaid taksonoomiamoduleid vastavalt kirjeldatud nõuetele;
- luua andmemudel, mis oleks piisavalt paindlik nõutava funktsionaalsuse katmiseks;
- luua eluslooduse klassifikatsioonide haldamise veebiteenus, kus oleks lahendatud olemasolevate süsteemide kitsaskohad ning mida kasutades saaks ehitada graafilise

¹<http://cybertaxonomy.eu/>

²<http://scratchpads.eu/>

³<http://www.globalnames.org/>

⁴http://en.wikipedia.org/wiki/Representational_state_transfer

haldusliidese.

Magistritöö on jaotatud 4 peatükki:

- Esimeses peatükis kirjeldatakse eluslooduse klassifikatsiooni mõistet. Selgitatakse taksonoomide tööd tuues välja taksonoomide peamised tegevused. Kirjeldatakse taksonitele nime andmise protsessi. Peatüki lõpus tuuakse välja töö käigus valmiva rakenduse loomise põhjused.
- Teises peatükis tuuakse välja kasutajapoolsed nõuded loodavale tarkvarale. Antakse ülevaade olemasolevatest rakendustest ning võrreldakse nõudeid olemasolevate rakenduste funktsionaalsusega.
- Kolmandas peatükis kirjeldatakse töö käigus loodud rakendust, selle loomise protsessi ning tehtud valikuid. Täpsemalt selgitatakse loodud andmemudelit ning REST stiilis veebiliidest. Viimaks tuuakse välja olulisemate funktsioonide käivitamiseks kuluv aeg erinevate suurustega klassifikatsioonide puhul.
- Viimases peatükis arutleb autor loodud taksonoomiamooduliga seotud edasise töö üle.

Lisas 1 on välja toodud bioloogiaalased terminid, mida antud magistritöös kasutatakse. Lisas 2 tuuakse välja antud töö käigus loodud veebiteenuse võimalikud sihtpunktid. Lisas 3 antakse loodud veebiteenuse erinevate sihtpunktide kõigi väljade lühikirjeldus. Lisas 4 esitatakse tabelina kõigi kasutuskiruse testide tulemused.

1 Taust

Käesolev peatükk annab ülevaate antud tööga seotud taustast ning seletab lahti bioloogiaalased terminid nagu taksonoomia, eluslooduse klassifikatsioon ja teised seotud mõisted. Alapeatükkides antakse ülevaade loodava rakenduse kasutajate tööst. Hilisemates alapeatükkides tuuakse välja käesoleva magistritöö tegemise motivatsioon, valmiva rakenduse kasutajad ning nende nõuded loodavale tarkvarale.

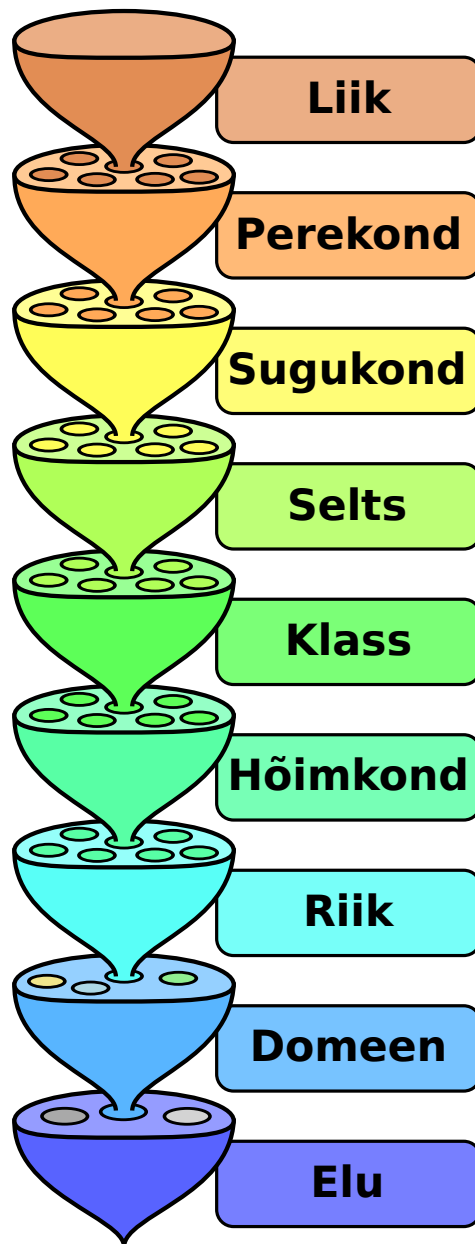
Taksonoomia on bioloogia haru, mis tegeleb eluta ja eluslooduse organismide klassifitseerimisega. Eluslooduse klassifikatsiooniks nimetatakse isendite grupeerimist taksonitesse vastavalt nende põlvnemisele⁵. Taksonid järjestatakse ühtseks süsteemiks, kus sarnased või seotud kategooriad (*taxonomic ranks*) on koondatud üldisemaks kõrgemaks kategooriaks. Üheks kategooriaks nimetatakse sarnaste tunnustega olemite kogumit [3]. Mida madalam taksoni kategooria, seda rohkem sarnasusi jagavad antud kategooriasse kuuluvad taksonid [4]. Joonisel 1 on välja toodud peamised taksonite kategooriad. Iga isend on paigutatud igasse kategooriasse olles seal mõne taksoni koosseisus.

Klassifitseerimise eesmärk on organismide paigutamine olemasolevasse klassifikatsiooni vastavalt sinna eelnevalt paigutatud organismide põlvnemisele ehk evolutsioonile. Näiteks paigutub taim, millel on eraldiseisvad lehed, juured, tüvi ning õied, kõige paremini taimede riiki (*Plantae*) katteseemnetaimede hõimkonda (*Magnoliophyta* või *Angiospermae*). Vastavalt detailsematele tunnustele paigutatakse käesolev taim järjest madalamatesse kategooriatesse, kuni liigini välja.

Lisaks peamistele taksonite kategooriatele kasutatakse tihtipeale ka vahepealseid kategooriaid. Näiteks võib liigi sees olla väiksemaid gruppe, kes omavahel erinevad. Neid gruppe kutsutakse alamliikideks.

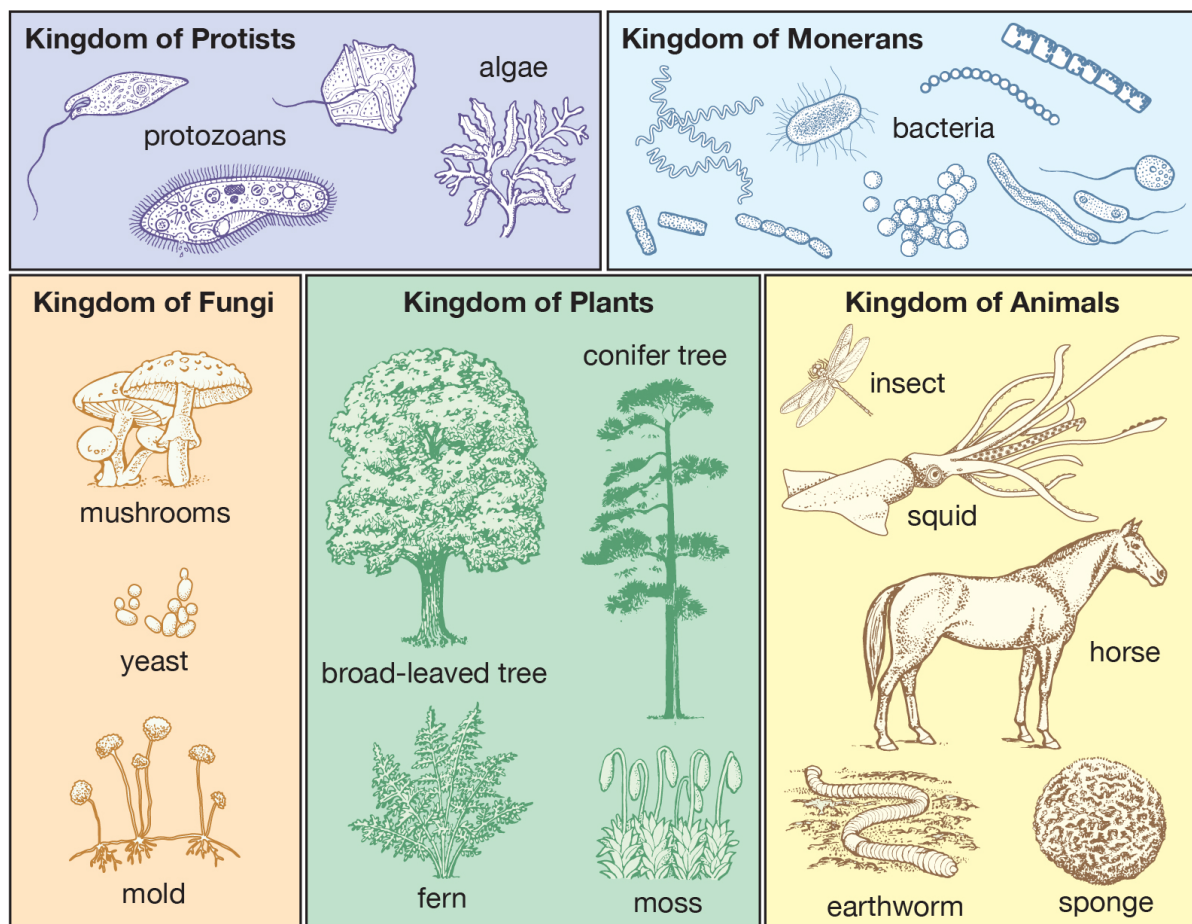
Bioloogiline klassifikatsioon loob korrastatud ja ühtse süsteemi nii olemasolevatele kui väljasurnud liikidele. Igale taksonile on antud teaduslik nimi. Taksonitele nime andmist reguleerivad rahvusvahelised nomenklatuuri koodeksid. Koodekseid standardiseerivad erinevate bioloogia harudega tegelevad rahvusvahelised asutused ja nõukogud. Näiteks võib

⁵Ökosüsteemide klassifikatsioone selles töös ei käsitleta aga põhimõtteliselt on ka nende puhul tegemist eluslooduse klassifikatsioonidega.



Joonis 1: Bioloogiline klassifikatsioon [5]

tuua loomade (ICZN)⁶, taimede ja seente (ICBN)⁷ ning bakterite (ICNB)⁸ rahvusvahelised nomenklatuurikoodeksid [6]. Antud standardiseerimine võimaldab bioloogidel üle maailma paremini tuvastada organisme. Lisaks standardiseeritud teaduslikule nimele on paljudel taksonitel ka tavanimetused, mis erinevad piirkonniti ja keelesti. Tavanimetuste rohkuse ja erinevuse tõttu on nad segadust tekitavad ning teadustöös mitesobivad. Näiteks ühte suurt hundi (*Canis lupus L.*) alamliiki teatakse erinevates paikades kui koer, *pini*, *dog*, *haushund*, *koira*. Sellel alamliigil on aga ainult üks teaduslik nimi - *Canis lupus familiaris*



© 2014 Encyclopædia Britannica, Inc.

Joonis 2: Viie riigi klassifikatsioon [7]

Klassifikatsioon annab võimaluse organiseerida fakte organismi gruppide ja üksikute organismide kohta. Näiteks hundid ja tiigrid kuuluvad kiskjaliste (*Carnivora*) seltsi. Samas hundid kuuluvad koerlaste (*Canidae*) sugukonda, mille enamikel liikmetel on pikad koonud ja pikakarvaline saba. Tiigrid aga kuuluvad kaslaste (*Felidae*) sugukonda, kelle liikmetele

⁶<http://iczn.org/code>

⁷<http://www.iapt-taxon.org/nomen/main.php>

⁸<http://www.ncbi.nlm.nih.gov/books/NBK8817/>

on iseloomulikud lühike koon ja lühikese karvaga saba.

Vastavalt Catalogue of Life⁹ iga-aastasele kokkuvõttele on 2014. aasta seisuga maailmas teadusele kirjeldatud üle 1.5 miljoni liigi [8]. Tänu teaduse ja tehnoloogiliste võimaluste arenemisele avastatakse ja klassifitseeritakse igal aastal tuhandeid uusi liike. Seetõttu võib uute andmete lisandumisel tekkida vajadus organisme ümber klassifitseerida ning taksonite kategooriaid ja nimesid täpsustada.

Taksonoomias alustatakse isendi klassifitseerimist tavaliselt tema liiki kuuluvuse määramisega. Kui sellist liiki ei ole veel teaduse poolt kirjeldatud, siis alustatakse liigi kirjeldamisega. Selle käigus tuleb otsustada millisesse perekonda see liik kuulub. Tavaliselt kaasneb sellega ka liigi kuuluvuse määramine kõrgematesse kategooriatesse kuni riigini. Evolutsioonilistes uurimustes seevastu võib näha vastupidist protsessi, kus eluslooduse süsteemi uurimise käigus jagatakse domeenid riikideks, need omakorda hõimkondadeks ja nii edasi. Eluslooduse klassifikatsioonide rohkuse tõttu ei ole üheselt määratud, mitmeks riigiks elusloodus jaguneb. Klassikaline, enam kui 50 aastat vana, riikideks jaotamine on välja toodud Joonisel 2.

Taksonite nimed ja klassifikatsioonid on tänapäeva bioloogias jätkuvalt peamiseks kommunikatsiooni vahendiks. Neid kasutavad kõik geneetika, ökoloogia, põllumajanduse, looduskaitse jne. andmebaasid. Näiteks ei ole üldjuhul võimalik talletada avalikes geenipankades geenijärjestusi, millel puudub taksoni teaduslik nimi [9]. Järelikult on väga oluline, et nimetatud valdkondade andmebaasid saaksid ligipääsu kvaliteetsetele klassifikatsioonidele ning nendega seotud tarkvaralistele töövahenditele. Antud magistritöö ülesandeks ongi luua tarkvaraline alus taksoni nimede ning nendel baseeruvate klassifikatsioonide haldamiseks.

1.1 Taksonoomia töö meetodid

Taksonoomilise töö meetodeid on erinevaid, kuid peamiselt on tegevused jagunenud järgnevalt:

1. Elusorganismide eksemplaride või nende kudede kogumine ning säilitamiseks ette

⁹<http://www.catalogueoflife.org/>

valmistamine või teistest kollektsioonidest eksemplari laenutamine. Samuti organismide uurimine nende looduslikes elukohtades.

2. Eksemplari võrdlev uurimine kasutades avaldatud töid ja kogudes talletatud isendeid.
3. Uuritava eksemplari määramine liigi tasemel. Uue liigi kirjeldamine juhul kui uurimistöö käigus selgub, et tegemist on siiani teadusele kirjeldamata taksoniga. Kirjeldavale taksonile korrektse nime valimine vastavalt rahvusvahelistele nomenklatuuri reeglitele.
4. Uuele taksonile evolutsiooniliselt parima asukoha määramine olemasolevas klassifikatsioonis. Uue avastusega kaasnenud klassifikatsiooni muudatuse ülevaatamise vajaduse kindlaks tegemine ning määramine.
5. Taksoni evolutsiooni välja pakkumine seoses tehtud avastusega.

Eeltoodud protsessi käigus kasutatakse erinevaid tunnuseid e. kriteeriume (sh. morfoloogia, füsioloogia, biokeemia ja geneetika). Eriti tormiliselt on viimase paarikümne aasta jooksul kasvanud geenijärjestuste kasutamine nii taksonite kirjeldamises kui ka nende evolutsioonilise päritolu uurimisel.

Eelpool kirjeldatud tegevuste eelduseks on üldtunnustatud taksonoomiliste kategooriate süsteem, nomenklatuurireeglite olemasolu ning protseduurid isendite määramiseks ja kontrollimiseks.

1.2 Taksonite nimetamine

Rahvusvaheliselt tunnustatud taksonite nomenklatuurireeglitenä kasutatakse Linnaean'i süsteemi millel andis esialgse kuju nüüdisaegse elusorganismide süstemaatika ja taksonoomia rajaja Carl von Linné. Nomenklatuurireeglid erinevad rühmiti, kuid sisaldavad alati järgnevaid elemente:

- taksoni nimi võib põhineda mistahes keelel, aga see tuleb vastavusse viia ladina keele grammatikaga;
- iga taksoni esimene korrektselt avaldatud teaduslik nimi on prioriteetne ning teised samale taksonile hiljem antud nimed on selle sünonüümid;

- ühe nomenklatuuri koodeksi piires loetakse korrektseks ainult unikaalseid nimesid - sama nime ei tohi anda erinevatele taksonitele¹⁰;
- taksoni nime kasutuse määrab viide tüüpeksemplarile.

Taksoni nimi on tänapäeva koodeksi reeglite põhjal alati seotud ainult ühe kindla eksemplariga, mis peab olema talletatud rahvusvahelist tunnustatud kogus. Sellist indiviidi nimetatakse tüüpeksemplariks. Kõigi teiste eksemplaride kuuluvus samasse liiki on aga teaduslik hüpotees [4, 11]. Taksoni nimega on seotud nomenklatuuristaatus, mis kirjeldab nime õigsust [12].

1.3 Taksonoomilised andmebaasid

Esimene taksonoomiline andmebaas märgiti ära 1972. aastal ilmunud artiklis, mille autoriteks olid Richard Swartz et al. Läbi aja on need andmebaasid oluliselt muutunud ning tänapäeval kasutatakse taksonoomilisi andmebaase järgnevalt:

- taksoninimestike loomiseks;
- tööks veebipõhiste liigikirjeldustega;
- bioloogiliste kogude haldamiseks (näiteks muuseumides, herbaariumides jm.) [13].

1.4 Klassifikatsioonide haldamine ja kasutamine

Klassifikatsioone talletatakse taksonoomilistes andmebaasides. Enamik klassifikatsioonide haldamiseks tehtud tööriistu toetavad vaid ühe klassifikatsiooni talletamist ning on loodud organisatsioonide poolt sisemiseks kasutamiseks, olles väljast poolt ligipääsmatud.

Leidub organisatsioone, kes koguvad erinevaid klassifikatsioone, püüdes neist luua ühte konsensusklassifikatsiooni. Tuntuim neist on Catalogue of Life, mis saab oma andmed rohkem kui 150 taksonoomilisest andmebaasist üle maailma [14]. Võimaldamaks eri klassifikatsioonides paiknevaid taksoneid omavahel ühendada, kasutab Catalogue of Life LSID'sid (Life Science Identifiers). Catalogue of Life'ga integreerimiseks tuleb ise leida

¹⁰Loomadel ja taimedel, seentel ja vetikatel on eri nomenklatuuri koodeksid. Loomade ja taimede taksoninimedel esineb palju kattuvusi ning kuna see tekitab segadust, siis on asutud looma ühist bioloogilist nomenklatuuri koodeksit BioCode [10].

tööriistad klassifikatsiooni loomiseks, kuna nende eesmärk ei ole mitte taksonoomilise tööriista pakkumine, vaid ajakohaste klassifikatsioonide kogumine. Samas leidub asutusi, kes arendavad ise taksonoomilisi klassifikatsioone (sh. peamised geenipangad).

Klassifikatsioonide kogujatele vastupidiselt leidub rohkem asutusi, kelle eesmärgiks on pigem klassifikatsioonide kasutamine. Tüüpiliselt kasutatakse selleks taksonoomiliste andmebaaside regulaarset alla laadimist. Näiteks võimaldab Catalogue of Life oma klassifikatsiooni alla laadida tervikuna või läbi veebiteenuse.

1.5 Motivatsioon

Eluslooduse klassifikatsioonide rohkuse tõttu on probleemiks erinevate klassifikatsioonide vaheline andmete haldamine. Vaja on bioloogiliste klassifikatsioonide haldamise süsteemi, mis võimaldaks talletada piiramatult erinevaid klassifikatsioone, sealhulgas erinevate klassifikatsioonide taksonite vahelisi seoseid ning lubaks teha toiminguid klassifikatsioonide sees ja vahel (näiteks taksonite ümberpaigutamine klassifikatsioonis, taksonite sidumine erinevate klassifikatsioonide vahel jms.).

Maailmas on loodud mitmeid tarkvarasid, mille eesmärgiks on taksonoomide töö lihtsustamine ning eluslooduse klassifikatsioonide haldamine. Samas on autorile teadaolevalt olemasolevates süsteemides puudusi (vt. peatükki Ülevaade rakendustest).

Tartu Ülikooli loodusmuuseum ja Eesti teaduse taristu teekaart NATARC¹¹ hoiab ning haldab Eesti peamist eluslooduse klassifikatsiooni kasutades selleks veebitöölauda PlutoF¹². Antud süsteemi eluslooduse klassifikatsiooni moodul võimaldab hallata korraga ühte klassifikatsiooni ning seab sellega kasutamiseks piirangud.

PlutoF veebitöölauda kasutatakse elurikkuse uuringuteks, seega selle süsteemi üks keskne osa on eluslooduse klassifikatsiooni moodul. Kogu elurikkusega seotud info peab alati olema seotud taksoni nimede ning nende asendiga klassifikatsioonis. Vastasel juhul on hoitava informatsiooni väärtus väga väike. Hallatavat eluslooduse klassifikatsiooni kasutavad ka kolmandad osapooled, näiteks Eesti Looduse Infosüsteem¹³, mis saab Tartu Ülikooli loodusmuuseumi käest perioodiliselt uuendatud klassifikatsiooni tabelarvutusprogrammi

¹¹<http://natarc.ut.ee>

¹²<https://plutof.ut.ee>

¹³<http://eelis.ee/>

failina.

Uute liikide kirjeldamiseks ning muus osas klassifikatsiooni arendamiseks saab praegust süsteemi kasutada alles töö lõppjärgus - klassifikatsiooni muudatuste sisse viimiseks. Vaja oleks võimalust talletada süsteemis mitut klassifikatsiooni, et toetada elurikkusega seotud teadustöö tegemist.

Taksonoomid ja bioloogid on spetsialiseerunud enamasti ühele organismirühmale, näiteks liblikaliste seltsile (*Lepidoptera*). Selleks, et teha ettepanekuid klassifikatsiooni muutmiseks antud alamrühma kontekstis, oleks vajalik käsil olev klassifikatsiooni alamosa teha privaatseks ning sõltumatuks üldisest klassifikatsioonist. Vajatakse võimalust muuta klassifikatsiooni ühte osa ilma tervikut muutmata.

Antud funktsionaalsus annab taksonoomidele võimaluse siduda oma muudatustega viiteid ning märkusi, samas kasutada oma privaatset klassifikatsioonipuud teadustöö kirjutamisel. Teadustöö avaldamisel saab taksonoom teha ettepaneku loodud alamklassifikatsioon sidumiseks tervikuga. Sidumist hõlpsustaks klassifikatsiooni alamosa kloonimisel loodud seosed alamklassifikatsiooni taksonite ja üldise klassifikatsiooni taksonite vahel.

1.6 Kasutajad

Loodava klassifikatsioonide haldamise süsteemi kasutajateks on PlutoF süsteemis oma andmebaase arendavad uurijad, asutused, harrastusteadlased, õpilased jt. Lisaks sellele leidub Eesti-põhiseid ja rahvusvahelisi portaale, mis kasutavad PlutoF süsteemis talletatavat klassifikatsiooni. PlutoF süsteemis olev klassifikatsiooni moodul asendatakse antud töö käigus valmiva eluslooduse klassifikatsioonide haldamise süsteemiga. Läbi PlutoF veebitöölaua on arendatava mooduli otsesed kasutajad:

- EELIS riigi infosüsteem, mis vajab hästi hallatavat, ajakohast ning muudatusi talletavat klassifikatsiooni omamaks informatsiooni Eestis paiknevate liikide kohta;
- väiksemad elurikkuse portaalid, näiteks eElurikkus¹⁴, mis on veebipõhine Eestis paiknevate liikide kohta infot kuvav portaal;
- erinevad asutused, mis pakuvad elurikkuse andmestikuga seotud teenuseid kogude andmete haldamiseks, loodusvaatluste sisestamiseks ja taksonoomilise teadustöö

¹⁴<http://elurikkus.ut.ee/>

tegemiseks. Näiteks Eesti Loodusmuuseumi ja Tartu Ülikooli loodusmuuseumi koostööna valminud Loodusheli mobiilirakendus¹⁵;

- Rootsi Riiklik loodusmuuseum¹⁶, kellega koostöös luuakse ühtset vabavaralist loodusteadeuslike kogude haldamise infosüsteemi DINA-WEB¹⁷.

Üldisemalt on loodava klassifikatsiooni haldamise süsteemi kasutajad erinevad portaaliid, mis koguvad, töötlevad, analüüsivad ja teevad kättesaadavaks elurikkuse andmeid.

1.7 Kasutajapoolsed nõuded

Kasutajapoolseteks nõueteks antud rakendusele käesoleva töö kontekstis on:

1. pakkuda veebiteenust nii andmete redigeerimiseks kui lugemiseks;
2. taksoninimede sünonüümideks tegemine;
3. hübriidide kirjeldamine;
4. nomenklatuuristaatuse lisamine;
5. taksoni nimega tehtavate muudatuste ehk aktide talletamine, sidudes need kasutajapoolsete tsiteeringute ja märkustega;
6. tervikliku klassifikatsiooni kloonimine;
7. klassifikatsiooni alamosa kloonimine;
8. uute klassifikatsioonide loomine;
9. erinevate klassifikatsioonide sidumine;
10. seoste küsimine (taksoni olemasolu teistes klassifikatsioonides);
11. tavanimetuste haldamine.

¹⁵<http://natmuseum.ut.ee/seminar>

¹⁶<http://www.nrm.se/>

¹⁷http://dina-project.net/wiki/Main_Page

2 Ülevaade rakendustest

Käesolev peatükk grupeerib kõik kättesaadavad eluslooduse klassifikatsioonidega tegelevad rakendused kolme rühma. Järgnevalt antakse ülevaade antud töö kontekstis olulistest rakendustest, tuues välja nii nende tugevad kui ka nõrgad küljed. Viimases alapeatükis võrreldakse eelnevalt välja toodud rakenduste vastavust antud magistritöös kirjeldatud kasutajapoolsetele nõuetele.

Eluslooduse klassifikatsioonide talletamisega tegelevad paljud asutused, kes on endale selle tarbeks loonud erineva funktsionaalsusega rakendusi. Need rakendused võib omaduste poolest jaotada kolme rühma - konsensusklassifikatsiooni arendamiseks loodud rakendused, ühe klassifikatsiooni arendamiseks loodud rakendused ning teadustöökäsitööks loodud rakendused.

Konsensusklassifikatsiooni loomiseks arendatud rakendused proovivad luua mitmete klassifikatsioonide sidumise läbi ühte terviklikku klassifikatsiooni. Seda klassifikatsiooni pakutakse läbi veebiteenuse kõigile kasutamiseks. Samas ei paku siia gruppi kuuluvad rakendused kasutajatele tarkvaralisi lahendusi antud süsteemis oma klassifikatsioonide loomiseks ja haldamiseks. Sellesse gruppi kuuluvad Encyclopedia of Life¹⁸ ja Catalogue of Life, mis teadustöökäsitööks vajaliku funktsionaalsuse vähesuse tõttu jäävad antud töö rakenduste võrdlusest välja.

Ühe klassifikatsiooni haldamiseks loodud andmebaase esineb mainitud rakendustest maailmas kõige enam. Neid platvorme arendavad organisatsioonid, kes annavad välja uusi liiginimesid, dokumenteerivad nende esmast kasutust ning tegelevad looduskaitsega. Lisaks haldavad need organisatsioonid ökoloogia, geneetika, evolutsiooni jt. andmebaase. Samas ei ole eesmärk pakkuda konsensusklassifikatsioone ega anda võimalust luua oma klassifikatsioone. Sellesse gruppi kuuluvad näiteks MycoBank¹⁹ ja Index Fungorum²⁰. Piiratud funktsionaalsuse tõttu neid andmebaase antud magistritöö raames lähemalt ei vaadelda.

Taksonoomiliseks teadustöökäsitööks loodud rakendused pakuvad kasutajatele võimalust luua, säilitada ja muuta eluslooduse klassifikatsioone. Nende rakenduste eesmärk on võimaldada klassifikatsioone arendavatel teadlastel oma tööd paremini teha. Siia gruppi kuuluvad Global Names Architecture, EDIT Platform for CyberTaxonomy ja Scratchpads biodi-

¹⁸<http://eol.org/>

¹⁹<http://www.mycobank.org/>

²⁰<http://indexfungorum.org>

versity online. Need rakendused sarnanevad funktsionaalsuse poolest antud magistritöös kirjeldatud nõuetega ning kuuluvad edasisse võrdlusesse.

Rakenduste hindamiseks ja omavahel võrdlemiseks kasutas autor taksonoomiliseks teadustööks loodud rakendusi. Rakenduste võrdlemiseks hinnati iga rakenduse vastavust kõigile nõuetele, mis on kirjeldatud peatükis Kasutajapoolsed nõuded. Iga nõude vastavuse hindamine toimus kolmepallisel skaalal, kus nõude mitte täitmine andis null punkti, nõude osaline täitmine andis ühe punkti ning nõude täielik täitmine andis 2 punkti.

2.1 Global Names Architecture

Global Names Architecture (GNA) on loodusteaduslik platvorm, mis võimaldab kasutajatel avastada, organiseerida ning ühendada loodusteaduslikke andmeid taksoni nimede kohta [15].

GNA kasutab taksonoomiamoodulina Hiina Teaduste Akadeemia poolt arendatud Taxonomic Tree Tool²¹ rakendust. Moodul võimaldab kasutajatel luua ja redigeerida taksonoomiaid. Süsteem võrdleb sisestatud puid leidmaks klassifikatsioonides nii vastuolulisi kui ka kattuvaid alasid.

Taxonomic Tree Tool on eraldiseisev veebipõhine rakendus, kus on võimalik luua ja talletada mitut klassifikatsiooni. Taksonite kohta saab talletada nime, autorit, kirjeldamise aastat, kategooriat ning sünonüümi nime, autorit ja kirjeldamise aastat. Autori arvates on üksikute taksonitega seotud andmete talletamine ning nende töötlemine liialt piiratud.

Tööriist on suunatud klassifikatsiooni, kui tervikuga töötamiseks. Märkusi ja tsiteeringuid saab lisada klassifikatsiooni, kuid mitte üksiku taksoni kohta. Võimalik on kloonida tervet klassifikatsiooni kui ka lihtsalt alamosa. Suurimaks tugevuseks on klassifikatsioonide vaheliste sarnasuste ja erinevuste automatiseeritud leidmine.

Peamisteks puudusteks Taxonomic Tree Tool juures on:

- veebiteenuse puudumine - andmetele pääseb ligi ainult läbi graafilise veebipõhise tööriista;
- taksoniga seotud andmete vähesus;

²¹<http://ttt.biodinfo.org/TF/>

- taksoniga tehtavate muudatuste ajaloo mitte talletamine;
- klassifikatsioonidevaheliste seoste puudumine;
- tavanimetuste haldamise puudumine.

2.2 EDIT Platform for CyberTaxonomy

Edit Platform for CyberTaxonomy on taksonoomilise töövoos kõiki aspekte kattev vabavaraliste tööriistade kogum. Platvorm võimaldab taksonoomide töös hallata järgnevaid alasid:

- taksonite redigeerimine;
- andmete avaldamine ja redigeerimine;
- andmete talletamine ja vahetamine;
- kollektsioonid ja isendid;
- kirjeldused;
- välitööd;
- kirjandusviited;
- geograafia.

Platvormi tuumaks on kogu taksonoomide tööd toetav andmemudel koos rakendustega enamike töö osade jaoks [16].

Edit Platform for CyberTaxonomy platvorm kasutab taksonoomiamoodulina allalaetavat kohalikus arvutis kasutamiseks mõeldud rakendustarkvara Taxonomic Editor²². Tarkvara tuleb siduda andmebaasiga, mis võib asuda nii kohalikus arvutis kui ka keskses serveris. Rakenduses on võimalik luua ja talletada mitmeid klassifikatsioone ning redigeerida taksonite andmeid.

Taxonomic Editor on suunatud taksonitega seotud andmete muutmisele ning võimaldab väga erinevate andmete salvestamist, mis on autori arvates antud süsteemi kõige suurem tugevus. Süsteem küll võimaldab mitmete klassifikatsioonide talletamist, kuid puudub võimalus teha toiminguid klassifikatsioonide vahel. Iga klassifikatsiooni käsitletakse kui

²²<http://cybertaxonomy.eu/taxeditor/>

kinnist hulka, mida ei ole võimalik siduda teiste klassifikatsioonidega. Samuti puudub võimalus klassifikatsiooni osaliselt ja täielikult kloonida.

Peamisteks puudustes Taxonomic Editor juures on:

- veebiteenuse puudumine - andmeid saab muuda ainult läbi kohalikus arvutis installitud tarkvara;
- klassifikatsioonidevaheliste tegevuste piiratus - ei ole võimalik klassifikatsioone kloonida ega taksoneid eri klassifikatsioonides omavahel siduda;
- taksonitega tehtavate muudatuste ajaloo mitte talletamine.

2.3 Scratchpads biodiversity online

Scratchpads on elurikkuse alast teadustööd toetav veebipõhine keskkond, mis lubab kõigil hallata, jagada ja avaldada oma andmeid ning luua oma teadusvõrgustikku. Antud keskkonnas saab luua elurikkuse lehekülgi geograafiliste regioonide kohta erinevatest taksonoomilistest gruppidest ning üldiselt erinevatest loodusteaduse aspektidest. Scratchpads'i peamisteks osadeks on tööriistad bioloogiliste klassifikatsioonide haldamiseks, kirjanduse haldamise, erinevate meediate haldamiseks (pildid, videod, helid, jne.) ning taksoni lehekülgede loomiseks (sisaldades taksoni kirjeldusi, isendite andmeid ning geograafilist paiknevust) [17].

Scratchpads'i taksonoomiamooduliga on võimalik importida olemasolevaid klassifikatsioone alamosana kui ka muuta ja täiendada juba olemasolevaid klassifikatsioone. Klassifikatsioonid on piiratud talletama taksoneid ühe riigi piires - tervikliku klassifikatsiooni haldamine ei ole toetatud. Iga klassifikatsiooni käsitletakse kui kinnist hulka, kus puudub võimalus siduda taksoneid erinevates klassifikatsioonides. Peamise tugevusena leidis autor antud taksonoomiamooduli juures võimaluse kirjeldada taksonite lisamisel ja muutmisel täidetavaid välju. Näiteks puudub esialgu võimalus lisada taksonite kohta nomenklatuuri-staatuseid, kuid selle välja saab mugavalt juurde lisada koos võimalike valikutega.

Peamised puudused Scratchpads biodiversity online juures on:

- veebiteenuse puudumine - andmeid saab muuta ainult läbi graafilise veebipõhise keskkonna;

- klassifikatsioonidevaheliste tegevuste piiratus - süsteemisisesid klassifikatsioone ei ole võimalik kloonida ega siduda teiste klassifikatsioonidega;
- klassifikatsioonide piiratus - klassifikatsioonid saavad olla maksimaalselt ühe riigi taksonite kohta. Tervikliku klassifikatsiooni talletamise võimalus puudub.

2.4 Tulemused

Rakenduste hindamise tulemused on välja toodud tabelis 1. Maksimaalne punktide arv on 22.

Tabel 1: Rakenduste võrdlus vastavalt kasutajapoolsetele nõuetele

Nõue	GNA	EDIT	Scratchpads
1. pakkuda veebiteenust nii andmete redigeerimiseks kui lugemiseks	0	0	0
2. taksoninimede sünonüümideks tegemine	2	2	2
3. hübriidide kirjeldamine	0	2	2
4. nomenklatuuristaatuse lisamine	0	2	2
5. taksoni nimega tehtavate muudatuste ehk aktide talletamine, sidudes need kasutajapoolsete tsiteeringute ja märkustega	0	0	2
6. tervikliku klassifikatsiooni kloonimine	2	0	0
7. klassifikatsiooni alamosa kloonimine	2	0	1
8. uute klassifikatsioonide loomine	2	2	1
9. erinevate klassifikatsioonide sidumine	0	0	0
10. seoste küsimine (taksoni olemasolu teistes klassifikatsioonides)	0	0	0
11. tavanimetuste haldamine	0	2	2
KOKKU	8	10	12

Kokku saab Global Names Architecture poolt kasutatav Taxonomic Tree Tool 8 punkti. Pakutav tööriist võimaldab iga taksoni kohta talletada vähesel määral andmeid (nimi, autor, sünonüümid). Tsiteeringuid ja märkusi saab lisada ainult klassifikatsioonide kohta, üksikute taksonite kohta ei ole võimalik lisada tsiteeringuid ega märkusi.

Edit Platform for Cybertaxonomy kasutatav taksonoomiamoodul saab kokku 10 punkti. Rakendus ei ole veebipõhine (nõuab kohalikku arvutisse paigaldamist) ning võimaldades küll mitut klassifikatsiooni talletada, puudub võimalus teha klassifikatsioonidevahelisi tehinguid (taksonite sidumine, kloonimine).

Scratchpads biodiversity online poolt kasutatav taksonoomiamoodul saab 12 punkti. Rakendusel on küll väga kasulik võimalus lisada uusi välju taksoni kirjetele, kuid puuduvad vahendid klassifikatsioonidevaheliseks tööks (klassifikatsiooni kloonimine, taksonite sidumine). Rakenduses puudub võimalus talletada terviklikku klassifikatsiooni.

3 Väljapakutud lahendus

Eeltoodud rakenduste võrdluse põhjal selgub, võrreldes olemasolevate rakenduste funktsionaalsust vajamineva funktsionaalsusega, et kõigil valitud taksonoomiamoodulitel on puudusi. Seega ei sobiks need rakendused Tartu Ülikooli loodusmuuseumi ja NATARC poolt arendatava PlutoF pilve taksonoomiamooduliks.

Käesoleva peatükk algab antud magistritöös loodud taksonoomiamooduli arendusprotsessi kirjeldusega. Seejärel antakse ülevaade väljapakutud andmemudelitest ning kirjeldatakse loodud REST stiilis veebiliidesest ehk REST API't (*Application programming interface*). Viimastes alapeatükkides testitakse rakenduse kiirust erineva suurusega klassifikatsioonide puhul ning tuuakse välja antud mooduli rakendusvõimalused.

3.1 Kasutatud tarkvaraarenduse võtted

Töö käigus loodud taksonoomiamooduli arendamiseks kasutati programmeerimiskeelt Python (versioon 2.7.5) koos veebiraamistikuga Django²³ (versioon 1.7.4). REST stiilis veebiliidese loomisel kasutati Django lisamoodulit Django REST Framework²⁴ (versioon 2.4.4), mis lihtsustas veebiteenuse funktsionaalsuse lisamist.

Rakendus on osaliselt kaetud automatiseeritud ühik- ning integratsioonitestidega. Testimisel kasutati Python'i standardset testimise raamistikku unittest²⁵.

Rakenduse andmebaasisüsteemiks valiti PostgreSQL²⁶, mille peamiseks eelisteks on suur kasutajaskond, laiendatavus ja standarditele vastavus [18]. PostgreSQL sai valituks, kuna seda kasutab sisemiselt ka PlutoF veebitöölaud, kuhu rakendus valmides integreeriti.

Antud veebiteenuse arendamine toimus iteratiivselt. Tagamaks kasutatava ning reaalsetele nõuetele vastava rakenduse valmimise, julgustab iteratiivne tarkvaraarendus kasutajatelt tagasisidet saama rakenduse arendamise protsessi käigus [19]. Iga iteratsiooniga lisati või muudeti rakenduse funktsionaalsust ning parandati leitud vead. Vajalik oli kasutada iteratiivset tarkvaraarenduse mudelit, kuna rakenduse loomise alguses olid ainult mõned nõuded paika pandud. Lõplik nõutav funktsionaalsus selgus mitmete iteratsioonide käigus.

²³<https://www.djangoproject.com/>

²⁴<http://www.django-rest-framework.org/>

²⁵<https://docs.python.org/2/library/unittest.html>

²⁶<http://www.postgresql.org/>

3.2 Andmemudel

Eluslooduse klassifikatsioon, nagu kõik klassifikatsioonid, on ülesehituselt puustruktuuriga - tippudel (taksonitel) võib olla nii vanemaid (kõrgema taseme taksonid) kui ka järglasi (madalama taseme taksonid). Puustruktuuri haldamine relatsioonilises andmebaasis muutub aga väga aeglaseks, kui puustruktuur kirjeldada läbi vanema viite, mis asub iga tipu juures. Nimetatud antimuster (*antipattern*) on tihtipeale esimene viis, kuidas proovitakse relatsioonilises andmebaasis puustruktuuri talletada. Selline andmete talletamine on mõistlik ainult siis, kui puu maksimaalne sügavus on kaks. Vastasel juhul muutuvad alampuu leidmise päringud väga aeglaseks [20].

Käesolevas töös loodava rakenduse üks enimkasutatavaid päringuid saab olema teadmata sügavusega alampuu pärimine. Ülalmainitud antimustri kasutamine muudaks rakenduse väga aeglaseks, kuna alampuu saamiseks tuleks läbi käia kõik tipud. Alternatiivseteks puu hoidmise mustriteks on: raja nummerdamise (*path enumeration*) muster, pesastatud hulkade (*nested sets*) muster ja sulundi tabeli (*closure table*) muster.

Raja nummerdamise mustri korral hoitakse iga puu tipu juures kogu teed alates kõige kõrgema taseme vanemast kuni tipu endani. Näiteks, kui tipu A otsene järglane on B ja tipu B otsene järglane on C, siis hoitakse tabelis radasid järgnevalt (kujul tipp-rada):

- A-“A/”
- B-“A/B/”
- C-“A/B/C/”

Antud mustri peamiseks probleemiks on raja haldamise tülikus ning andmebaasipoolse valideerimise puudumine - rada hoitaks baasis teksti tüüpi väljas.

Pesastatud hulkade mustriga hoitakse iga tipu juures kahte numbrit - **phvasak** ja **phparem** väärtus. Neid numbreid antakse tippude järgnevalt: **phvasak** on väiksem kui ühegi teise alamtippu **phvasak** väärtus, **phparem** on suurem kui ühegi alamtipu **phparem** väärtus. Need numbrid ei ole seotud tippude teiste väärtustega, kasutatakse ainult puustruktuuri hoidmiseks. Kasutades ülaltoodud näidet tippude A, B ja C kohta, on pesastatud hulkade korral tabelis read (kujul tipp-phvasak-phparem):

- A-1-6

- B-2-3
- C-4-5

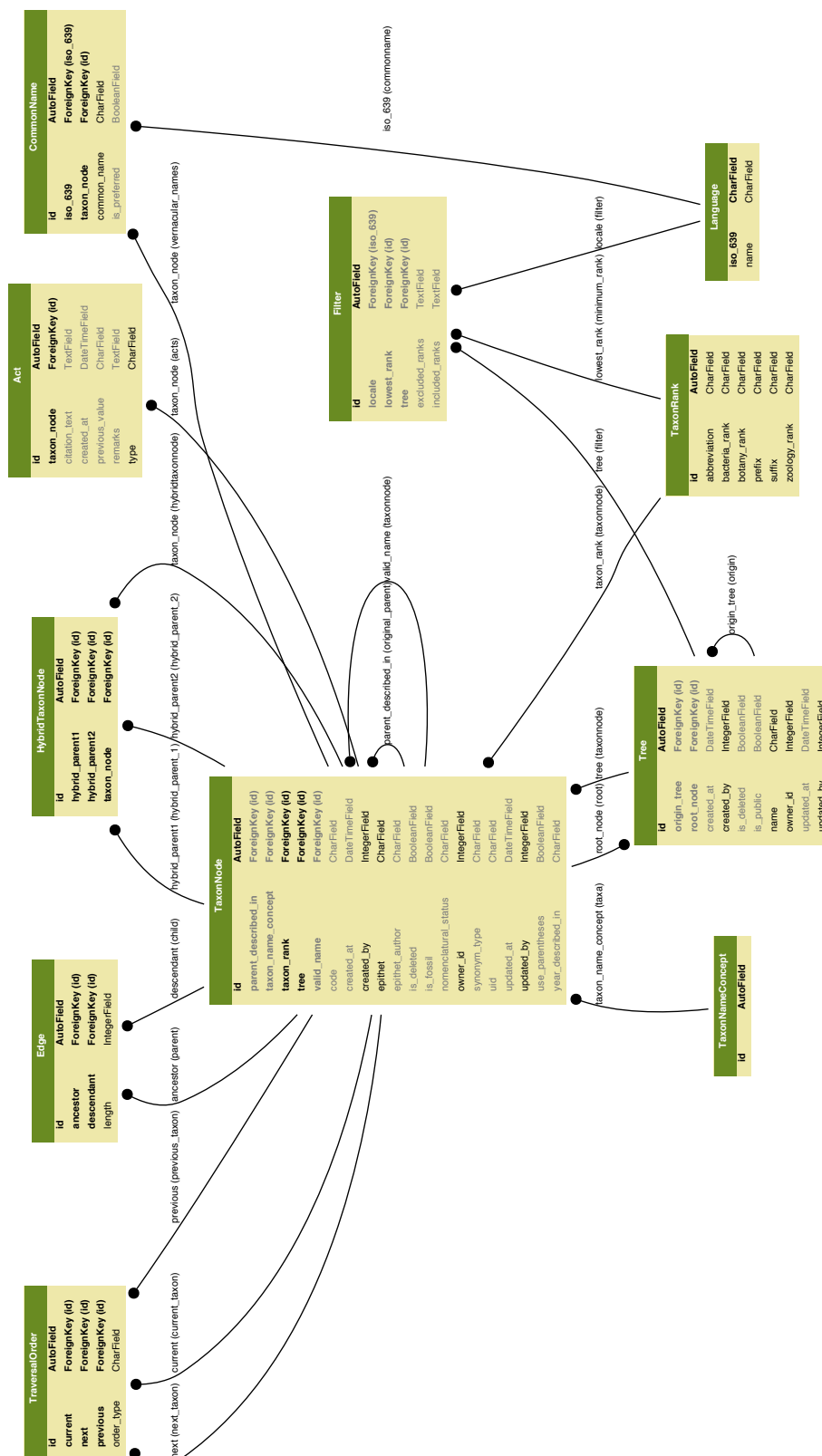
Selle mustri peamine tugevus on see, et tipu kustutamisel jääb puustruktuur samaks. Samas on otsese vanema ja otsese järglase pärimine tunduvalt keerukam kui eelnevate võimaluste puhul. Puustruktuuri muutuste puhul ei peeta pesastatud hulkade mustrit kõige paremaks, kuna see nõuab olulist **phvasak** ja **phparem** väärtuste ümberarvutamist [20].

Sulundi tabeli muster on ainus, mille juures kasutatakse puustruktuuri hoidmiseks välist tabelit. Antud välises tabelis on kirjas kõik vanem-järglane-kaugus seosed, mitte ainult otsesed vanem-järglane-kaugus seosed. Kauguseks nimetatakse sulundi tabelis kahe tipu vahelist kaugust. Kasutades raja nummerdamise juures toodud tippude A, B ja C näidet, on sulundi tabelis kirjas read (kujul vanem-järglane-kaugus):

- A-A-0
- A-B-1
- B-B-0
- A-C-2
- B-C-1
- C-C-0

Antud mustri kasutamisel on väga mugav pärida alampuud ning otseseid vanemaid ja järglasi. Samuti on mugav taksonit kustutada ning alampuud ümber paigutada. Puuduseks võib antud mustri juure tuua selle, et sügavate ja suurte puude korral muutub sulundi tabel väga mahukaks.

Sulundi tabeli muster on arusaadavam kui pesastatud hulkade muster. Mõlemal on omad eelised, kuid sulundi tabeli mustri korral on hierarhiat lihtsam hallata. Lisaks on sulundi tabeli mustri korral alampuudega tehtavad tehingud (alampuu paigutamine teise harusse, tippude lisamine ja kustutamine) lihtsamini teostatavad [20].



Joonis 3: Andmemudel

Magistritöös loodava rakenduse arendamiseks kirjeldati paindlik andmemudel, mille abil oleks võimalik täita kõik kasutajapoolsed nõuded. Alapeatüki alguses toodu põhjal on antud andmemudelis klassifikatsiooni talletamise jaoks otsustatud sulundi tabeli mustri rakendamise kasuks. Loodud andmemudel on kujutatud joonisel 3. Tabelis 2 on välja toodud taksonoomiamoodulisse kuuluvad olemid koos lühikese kirjeldusega.

Tabel 2: Andmemudeli olemite kirjeldus

Olemi nimetus	Olemi kirjeldus
TaxonNode	hoiab andmeid klassifikatsiooni kuuluvate taksonite kohta
Edge	hoiab sulundi tabeli mustri järgi klassifikatsioonipuude struktuuri
TaxonNameConcept	seob erinevates klassifikatsioonides paiknevaid taksoneid
Tree	hoiab andmeid klassifikatsioonide kohta
TaxonRank	kirjeldab taksonoomilisi kategooriaid
Filter	võimaldab erinevaid parameetreid kasutades kitsendada taksonite otsingut
CommonName	hoiab taksonite tavanimetusi
Language	võimaldab tavanimetuste hoidmisel valida konkreetse tavanime- tuse keelt
HybridTaxonNode	võimaldab hoida hübriidide kohta infot
TraversalOrder	võimaldab hoida erinevaid klassifikatsioonipuu läbimise järjekor- di
Act	hoiab taksonitega seotud tähtsamate muudatuste ajalugu

3.3 Veebiteenus

Veebiteenuseks nimetatakse tarkvarakomponenti, mille poole teised rakendused pöörduvad üle interneti [21, 22]. Magistritöös loodud tarkvara veebiliides toetab suhtlust kasutades JSON andmevahetusformaati²⁷.

Loodud rakendus on arendatud kasutades Django veebiraamistikku. Django põhineb mudel-vaade-kontroller arhitektuurimustril. Antud taksonoomiamoodul kasutab Django

²⁷<http://www.json.org/>

poolt pakutavat rakenduse ülesehitust (seaded, andmebaasiga suhtlus ja muu vajalik rakenduse käivitamiseks) ning olemite kirjeldamise funktsionaalsust. Vaikimisi ei ole Django raamistikul REST veebiteenuse loomist lihtsustavaid vahendeid, seetõttu kasutati vabavaralist Django REST raamistikku, mis on ka ühtlasi kõige levinum viis Django veebiraamistikus REST veebiteenuse loomiseks [23]. Lisaks heale ülesehitusele on REST raamistiku üks parimaid omadusi võimalus luua ennastdokumenteeriv ja veebis lehitsetav veebiteenus [24]. Django REST raamistikuga asendati Django vaated ja kontrollid veebiteenuse loomiseks mõeldud spetsiaalsete vaadete ja kontrollitega.

Joonisel 4 on toodud ekraanipilt lehitsetava veebiteenuse ühest leheküljest. Veebilehel on võimalik vaadata iga sihtpunkti (*endpoint*) aktsepteeritavaid parameetreid, sirvida kättesaadavaid olemeid ning neid kustutada. Lehitsetava veebiteenuse kasutajal on võimalik uusi andmeid lisada ning neid muuta iga olemi (ainult muutmine) ja nimistu (ainult lisamine) vaates. Veebiteenus on mõeldud teiste süsteemidega integreerimiseks, seega saab kõik päringud teha ka otse universaalse ressursi lokaatori (URL) sihtpunktide vastu. Selleks tuleb kasutada HTTP protokollis kirjeldatud meetodeid GET, POST, PUT, DELETE.

Kõigi tavanimetuste pärimiseks süsteemist tuleb jooksutada käsk “`curl -X GET http://localhost:8000/api/taxonomy/vernacular_name/`”, mis tagastab JSON tüüpi objekti, mida on näha joonisel 4.

Ühe ressursi nimekirj tagastatakse kui saadetakse GET päring aadressile `/api/taxonomy/<ressursi nimi>/`. Üksiku ressursi pärimiseks tuleb vastavalt REST põhimõtetele saata GET päring aadressile `/api/taxonomy/<ressursi nimi>/<olemi id>/`. Sarnaselt tuleb uue ressursi loomiseks saata POST päring aadressile `/api/taxonomy/<ressursi nimi>/` ning olemasoleva ressursi muutmiseks tuleb saata PUT päring aadressile `/api/taxonomy/<ressursi nimi>/<olemi id>/`.

Kõigi järjendit tagastavate sihtpunktide korral on kasutatud andmete lehekülgedeks jaotamist (*pagination*). Vaikimisi tagastatakse tulemusi 10 vaste kaupa. Võimalik on päringule kaasa anda parameetreid tagastatavate andmete alguspunkti (*offset*) ja limiidi (*limit*) kohta. Näiteks, kui järjendi päringu lõppu kirjutada `?offset=5&limit=50`, siis tagastatakse 50 järgmist taksonit alates 5. päringu tulemusest. Eelnimetatud parameetritele lisaks saab tagastatava hulga teist lehekülge pärida parameetriga `?page=2`.

Lisas 2 on välja toodud kõik loodud veebiteenuse sihtpunktid koos kirjeldustega. Lisas 3 on

← → ↻ 🔍 ☆ 🛑 💬 ☰

Django REST framework v 2.4.4 siim ▾

Api Root ▸ Vernacular Name List

Vernacular Name List OPTIONS GET ▾

GET /api/taxonomy/vernacular_name/

HTTP 200 OK
Content-Type: application/json
Vary: Accept
Allow: GET, POST, HEAD, OPTIONS

```
{
  "count": 1,
  "next": null,
  "previous": null,
  "results": [
    {
      "id": 1,
      "url": "http://localhost:8000/api/taxonomy/vernacular_name/1/",
      "iso_639": "http://localhost:8000/api/taxonomy/language/est/",
      "common_name": "Taimed",
      "taxon_node": "http://localhost:8000/api/taxonomy/taxon/10/",
      "is_preferred": true
    }
  ]
}
```

Raw data HTML form

iso 639:

Language

common name:

Taxon node:

is preferred: ☐

Joonis 4: Lehitsetava veebiteenuse ekraanipilt

antud lühikirjeldus iga sihtpunktiga seotud väljade kohta. Veebiteenusega integreerimiseks detailsem dokumentatsioon, mis kirjeldab erinevate meetodite võimalikke parameetreid, on antud rakenduses endas sirvitava veebiteenusena. Näiteks joonisel 4 näidatud sinist nuppu “OPTIONS” vajutades saab näha kõiki võimalikke HTTP meetodeid koos iga välja tüübiga, mida see sihtpunkt võimaldab. Järgnevad alapeatükid annavad ülevaate enimkasutatud mittetriviaalsetest sihtpunktidest.

3.3.1 Aktide filtreerimine

Kõiki taksonitega teostatud olulisi muudatusi ehk akte on võimalik filtreerida taksoni identifikaatori (“taxon_node”), klassifikatsiooni identifikaatori (“tree”) ja tüübi (“type”) järgi. Lisaks on võimalik pärida kõiki muudatusi, mis on toimunud pärast kirjeldatud aega (“min_datetime”) või siis enne kirjeldatud aega (“max_datetime”). Samuti on võimalik antud otsingu parameetreid kombineerida. Näiteks GET päring `/api/taxonomy/act/?type=new_taxon&min_datetime=2015-01-31%2000:00:00` tagastab kõik taksoni lisamise aktid, mis on salvestatud alates 31. jaanuari 2015.

3.3.2 Alamklassifikatsiooni pärimine

Alamklassifikatsiooni pärimine tagastab etteantud taksoni kogu alamklassifikatsiooni. Näiteks kasutatakse seda sihtpunkti, kui soovitakse parema ülevaate andmiseks kuvada konkreetse taksoni mitme järgneva taseme järglased. Tagastatavaid tasemeid on võimalik määrata kasutades filtreid (allpool). Samuti võimaldab selline sihtpunkt teostada kindla taksoni alamklassifikatsiooni andmete eksporti. Alamklassifikatsioon on seotud ühe taksoniga, seega REST põhimõtete järgi [25] on alamkollektsioon kättesaadavaks GET tüüpi päringuga kasutades URL’i `/api/taxonomy/taxon/<taksoni id>/subtree/`.

3.3.3 Kõigi vanemate pärimine

Taksoni kõiki vanemaid päritakse selleks, et kuvada taksoni paiknevust klassifikatsioonis alates kõige kõrgema kategooria taksonist. Seda sihtpunkti kasutavad nii taksonoomid oma teadustöös ühe taksoni kuvamisel kui ka elurikkuse portaalid, mis kuvavad liikidega seotud infot. Antud ressurss on kättesaadav GET päringuga kasutades URL’i `/api/taxonomy/taxon/<taksoni id>/higher_taxa/`.

3.3.4 Kattuv klassifikatsiooni tee

Etteantud taksonite kattuva klassifikatsiooni tee päring tagastab antud taksonite hierarhiliste teede kõik kattuvad taksonid. Päringut kasutatakse leidmaks antud taksonite evolutsioonilist kaugust üksteisest. Vanemate ühisosa teada saamiseks tuleb saata GET päring URL'le `/api/taxonomy/taxon/<taksoni id>/higher_taxa_intersection/?given_ids=<taksoni id>,<taksoni id>,....`

3.3.5 Otseste järglaste pärimine

Otseste järglaste pärimine tagastab järjendi kõigist ette antud taksoni otsestest alamtaksonitest käesolevas klassifikatsioonis. Seda sihtpunkti kasutatakse tihti klassifikatsiooni graafilise esitlemise korral. Antud ressurss on kättesaadav GET päringuga URL'le `/api/taxonomy/taxon/<taksoni id>/direct_children/`. Kogu klassifikatsiooni esimese taseme taksonite pärimiseks tuleb saata GET päring URL'le `/api/taxonomy/tree/<klassifikatsiooni id>/root_children/`.

3.3.6 Taksonite filtreerimine

Taksoneid saab filtreerida kasutades taaskasutatavaid filtreid. Filtrite loomisel tuleb kasutada minimaalselt ühte tabelis 3 asuvatest filtreerimisvõimalustest. Filtri saab kaasa anda iga võimaliku taksonite järjestuse päringuga. Näiteks saab alamklassifikatsiooni järjestust filtreerida GET päringuga URL'le `/api/taxonomy/taxon/<taksoni id>/subtree/?filter=<filtri id>`.

Tabel 3: Filtrite koostamise võimalikud valikud

Parameeter	Tagastatavad taksonite omadused
locale	tavanimetuse olemasolu filtriga kaasa antud keeles
lowest_rank	kategooria on kõrgem kui etteantud kategooria
included_ranks	kategooria on üks etteantud kategooriatest
excluded_ranks	kategooria ei ole etteantud kategooriate listis
tree	kuuluvus etteantud klassifikatsiooni

3.3.7 Taksonite otsimine

Taksonite otsimiseks kasutatakse täisteksti otsingutarkvara Elasticsearch²⁸. Loodud taksonoomiamoodulis toetatakse taksoni otsimist teadusliku nime järgi tehes GET päringu URL'le `/api/taxonomy/taxon/search/?search_query=<märksõna>`. Lisaks on võimalik otsida tavanimetusi tehes GET päringu URL'le `/api/taxonomy/vernacular_name/search/?search_query=<märksõna>`.

3.4 Automatiseeritud taustategevused

Eelnevates peatükkides kirjeldati veebiteenust läbi REST liidese kasutatava funktsionaalsuse. Lisaks eeltoodule teostab rakendus erinevate tegevuste korral mitmeid taustategevusi.

Tabel 4: Aktide kirjeldused

Akti tüüp	Kirjeldus
<code>new_taxon</code>	Lisati uus takson
<code>marked_as_synonym</code>	Takson märgiti sünonüümiks
<code>marked_as_basionym</code>	Takson märgiti basionüümiks
<code>marked_as_current</code>	Takson märgiti kehtivaks nimeks
<code>change_parent</code>	Takson muutis asukohta klassifikatsioonis
<code>edit_name</code>	Taksoni nime muudeti
<code>change_nomen_status</code>	Taksoni nomenklatuuristaatust muudeti
<code>delete_taxon</code>	Takson kustutati

Taksonid on rakenduses kesksel kohal, seega on vajalik salvestada nendega toimuvate muutuste ajalugu. Selleks kasutatakse antud taksonoomiamoodulis akte. Taksoni mingi olulise väärtuse muutmise korral salvestatakse akte automaatselt. Kõik olulised muudatused on kirjeldatud tabelis 4.

Iga klassifikatsiooni on võimalik läbida eesjärjestuses (*pre-order*) puu läbimise algoritmiga. Antud funktsionaalsust kasutatakse mõnes elurikkuse andmeid kuvavas portaalis klassifikatsioonipuu läbimiseks. Selleks luuakse iga taksoni kohta üks `TraversalOrder`

²⁸<https://www.elastic.co/products/elasticsearch>

objekt, milles on määratud taksonile eelnev ja järgnev takson. Antud objekti uuendatakse klassifikatsioonipuu struktuuri muutuse korral.

Iga kloonitud klassifikatsiooni takson on automaatselt seotud esialgses klassifikatsioonis oleva taksoniga läbi olemi `TaxonNameConcept`. Taksoneid on võimalik siduda ka manuaalselt teistes klassifikatsioonides olevate taksonitega. Selleks tuleb taksoni muutmisel anda kaasa `TaxonNameConcept` väärtus, mis võib kuuluda mõnele teise klassifikatsiooni taksonile.

3.5 Kasutuskiirus

Rakenduse enimkasutatud tegevuste kiiruse mõõtmiseks kirjutas autor programmeerimiskeeles Python andmegeneraatori, millega täideti etteantud klassifikatsioon soovitud hulga taksonitega. Seejärel mõõdeti erinevate klassifikatsioonidega seotud toimingutele kulunud aega, kasutades programmi `cURL`²⁹. Erinevate toimingute soorituskiirust mõõdeti kolme erineva suuruse klassifikatsiooniga selleks, et vaadata kuidas rakenduse kiirus muutub erinevate andmehulkade korral. Iga kombinatsiooni toimingutest ja andmehulkadest mõõdeti viiel korral. Seejärel arvutati iga eelnevalt mainitud kombinatsiooni keskmine aeg. Erinevate mõõtmistulemuste algandmed on esitatud lisas 4, keskmised ajad on toodud tabelis 5.

Mõõtmiste ajal ei kasutatud arvutit ressursinõudlike tegevuste jaoks. Mõõtmised teostati arvutiga, millel on järgmised parameetrid:

- protsessor: 2.8 GHz Intel Core i7, 2 tuuma
- mälu: 16GB 1600 MHz DDR3
- kõvaketas: 500 GB SSD-tüüpi
- operatsioonisüsteem - OS X 10.9.5

Tabelist 5 selgub, et testitud päringute korral ei kasva päringu tegemiseks kuluv aeg oluliselt, kui taksonite arv klassifikatsioonis suureneb. Erandiks on klassifikatsiooni alamosa ümberpaigutamine hierarhias ning kloonimine. Nimetatud tegevuste korral kasvab päringu tegemiseks kuluv aeg lineaarselt tehinguga seotud taksonite arvuga. Seega kulub näiteks 1000 taksoniga klassifikatsiooni ümberpaigutamisele keskmiselt 15 sekundit. 1000 taksoniga

²⁹<http://curl.haxx.se/>

Tabel 5: Rakenduse kasutuskiirus erineva suurusega klassifikatsioonidega

Tegevus / Taksonite arv	1 000	10 000	100 000
Kogu klassifikatsiooni pärimine (10 esimest vastet)	0.25s	0.24s	0.29s
Kogu klassifikatsioon pärimine (lehtedeks jaotamata)	1.72s	1.77s	1.75s
Klassifikatsiooni alamosa pärimine (10 esimest vastet)	0.23s	0.23s	0.23s
Otsingu teostamine (10 esimest vastet)	0.10s	0.10s	0.11s
Kõigi vanemate pärimine (10 esimest vastet)	0.07s	0.08s	0.08s
Kõigi vanemate ühisosa (10 esimest vastet)	0.07s	0.09s	0.09s
Otseste laste pärimine (10 esimest vastet)	0.15	0.15s	0.15s
Taksonite filtreerimine keele põhjal (10 esimest vastet)	0.19s	0.19s	0.21s
Taksonite filtreerimine madalama kategooria põhjal (10 esimest vastet)	0.18s	0.18s	0.18s
Taksonite filtreerimine sisalduvate kategooriate põhjal (10 esimest vastet)	0.18s	0.18s	0.18s
Taksonite filtreerimine välja jäetud kategooriate põhjal (10 esimest vastet)	0.17s	0.19	0.22s
Taksonite filtreerimine klassifikatsiooni põhjal (10 esimest vastet)	0.18s	0.18s	0.20s
Klassifikatsiooni kloonimine (ühe taksoni keskmine aeg)	0.050s	0.050s	0.050s
Klassifikatsiooni alamosa ümber paigutamine (ühe taksoni keskmine aeg)	0.015s	0.015s	0.015s

alamosa kloonimisele kulub ligikaudu 50 sekundit. Need päringud on mainitud ajakuluga, kuna iga klassifikatsiooni kloonimisel tuleb sisuliselt luua uus klassifikatsioon ning alamosa ümberpaigutamisel tuleb ümber arvutada kogu alamosa. Antud tegevused on suhteliselt harva esinevad, seega on aktsepteeritav, et tegevustele kuluv aeg on pikk.

3.6 Rakendusvõimalused

Magistritöö jooksul loodud rakendus on kasutusel Tartu Ülikooli loodusmuuseumi poolt arendatavas PlutoF pilves. Joonisel 5 on näidatud PlutoF pilvega integreeritud taksonoomiamooduli graafiline liides, mis taustarakendusena (*backend*) kasutab antud töös

arendatud taksonoomiamoodulit.

PlutoF pilves kasutavad antud taksonoomiamoodulit läbi graafilise kasutajaliidese asutused, kes hoiavad selles süsteemis oma loodusteaduslike kogude andmeid. Nendeks asutusteks on Tallina Botaanikaaed, Eesti Maaülikool, Tartu Ülikool ning Eesti Loodusmuuseum. PlutoF kasutajatest on 58% Eesti-sisesed ning ülejäänud välismaised. PlutoF pilve kaudu kasutavad magistritöös arendatud taksonoomiamoodulit REST veebiliidesena Loodusheli mobiilirakendus ning 5D Visioon OÜ poolt arendatavad loodusteaduslikud rakendused. Lisaks võib ära mainida ka EELIS riigi infosüsteemi, mis kasutab enda klassifikatsiooni ajakohasena hoidmiseks PlutoF taksonoomiamooduli veebiteenust, olles lokaalselt klooninud klassifikatsiooni ning pärides iga teatud aja tagant aktide kaudu muudatuste infot.

2015. aasta 20. aprilli seisuga hoitakse PlutoF veebitöölaual enam kui 222 000 taksonit kokku kahes erinevat klassifikatsioonis.

Lisaks eelnevalt mainitud rakendusviisidele kuulub arendatud taksonoomiamoodul üleeuroopalise projekti DINA³⁰ pakutavate loodusteaduslike moodulite hulka. DINA eesmärgiks on pakkuda laiaulatuslikku veebipõhist loodusteaduslikku infosüsteemi, mis on suunatud kogude haldajatele.

³⁰http://dina-project.net/wiki/Main_Page

Classifications
New
Info

Search by taxon name

Q

← Back

PlutoF main tree

Edit access

Properties

History

Full taxon name	Agaricomycetes
Taxon name	Agaricomycetes
User defined ID	170609
Uid	
Epithet	Agaricomycetes
Author	
Taxon rank	Class
Year described in	
Valid name	
Synonym type	
Nomenclatural status	
Is fossil	false
Common name	

- + KGD Plantae
- + KGD Animalia
- KGD Fungi
 - Add child
 - + PHY Ascomycota
 - PHY Basidiomycota
 - Add child
 - SBP Agaricomycotina
 - Add child
 - + CLS Agaricomycetes
 - + CLS Dacrymycetes
 - + CLS Incertae sedis
 - + CLS Tremellomycetes
 - + SBP Incertae sedis
 - + SBP Pucciniomycotina
 - + SBP Ustilaginomycotina
 - + PHY Blastocladiomycota
 - + PHY Chytridiomycota
 - + PHY Glomeromycota
 - + PHY Incertae sedis
 - + PHY Neocallimastigomycota
 - + PHY Rozellomycota
 - + PHY Zygomycota
 - KGD Protozoa
 - + KGD Chromista
 - + KGD Bacteria
 - + KGD Protista

Start a new taxon tree

Edit

Joonis 5: Taksonoomiamooduli graafiline kasutajaliides PlutoF veebitöölal

4 Edasine töö

Käesolev peatükk annab ülevaate loodud taksonoomiamooduli potentsiaalsetest edasistest arengutest. Autor toob välja lisafunktsionaalsust ning muus osas süsteemi täiustamiseks tehtavaid muudatusi, mis võiks muuta moodulit enam kasutatavaks. Loodud taksonoomiamoodul on mõeldud kasutamiseks komponendina suuremas loodusteaduslikus süsteemis. Antud peatükis kirjeldatakse mooduli võimalikku edasist arengut.

Autori arvates oleks kasulik arendada võimalus andmete failist importimiseks ja eksportimiseks kasutades üldtuntud standardit taksonoomiliste andmete vahetamiseks - Taxonomic Concept Transfer Schema³¹. Tekib võimalus teistest süsteemidest klassifikatsioone ekspordida ning kasutada läbi loodud taksonoomiamooduli ilma importimiseks vajalikku skripti loomata.

Sarnaselt Global Names Architecture poolt kasutatavale taksonoomiamoodulile tuleks arendada võimalus klassifikatsioonidevaheliste sarnasuste leidmiseks. Antud funktsionaalsust võiks aga laiendada, sidudes samasisulised taksonid erinevates klassifikatsioonides automaatselt.

Ajaliselt kulukate tegevuste korral peaks olema võimalik küsida ajahinnangut töö valmis saamise kohta. Näiteks võiks klassifikatsiooni kloonimise ja taksonite klassifikatsioonis ümberpaigutamise kohta olla võimalik küsida eeldatavat lõpuaega.

Kloonitud klassifikatsioonis tehtud muudatusi ei ole hetkel võimalik automatiseeritult ühildada kloonimise aluseks olnud klassifikatsiooniga. Moodulisse tuleks lisada sihtpunktid, mis võimaldaks nimetatud tegevuse läbiviimist (sealhulgas erisuste lahendamised).

Traditsiooniliselt on kasutatud relatsioonilisi andmebaase hierarhiliste andmete hoidmiseks ning selleks on loodud ka mitmeid andmebaasi mustreid [20]. Siiski tasuks kaaluda klassifikatsioonide üleviimist mitterelatsioonilisse andmebaasisüsteemi (näiteks graafi andmebaasisüsteemi), kuna nad sobivad ülesehituse poolest kõige paremini hierarhiliste andmete talletamiseks [26]. Seda tuleks teha ainult siis, kui on loodud lihtne prototüüp-rakendus ning see on tõestanud olulisi eeliseid.

³¹<http://www.tdwg.org/standards/117/>

Kokkuvõte

Käesolevas töös käsitleti elurikkusealaseid klassifikatsioone haldavaid tarkvarasid ning arendati veebiteenus bioloogiliste taksonoomiate paremaks haldamiseks. Töö käigus täideti kõik püstitatud eesmärgid.

Magistritöö esimeses osas selgitati elurikkuse klassifikatsiooni mõistet ning sellega seonduvaid termineid. Kirjeldati taksonoomide tööd ning toodi välja kasutajapoolsed nõuded Tartu Ülikooli loodusmuuseumi poolt vajaminevale taksonoomiamoodulile. Selgus, et vajatakse klassifikatsioonide haldamise veebiteenust, mida oleks võimalik integreerida asutuse poolt arendatavale elurikkuse veebitöölauale.

Teises osas uuriti elurikkusealaseks teadustööks loodud klassifikatsioonide haldamise süsteeme. Võrreldi neid püstitatud nõuetega ning leiti, et kõigis olemasolevates taksonoomiamoodulites leidis puudusi.

Magistritöö kolmandas osas kirjeldati programmeerimiskeeles Python loodud REST veebiliidese arendusprotsessi ja tulemusi. Võrreldi erinevaid võimalusi hierarhiliste andmete talletamiseks relatsioonilises andmebaasis, mille tulemusena leiti, et muutuva iseloomuga eluslooduse klassifikatsioonide haldamiseks sobib kõige paremini sulundi tabeli (*closure table*) andmebaasi muster. Järgnevalt kirjeldati loodud tarkvara arendusprotsessi ning vaadeldi veebiteenuse enimkasutatud sihtpunkte. Peatüki lõpus testiti taksonoomiamooduli kasutuskiirust ning anti ülevaade rakendusvõimalustest. Selgus, et loodud veebiteenuse kiirusnäitajad on head.

Autorile tundus, et lisaks kasutajapoolsetele nõuetele oleks võimalik loodud rakendust veel täiendada. Aluseks võiks võtta võrdluses olnud süsteemide tugevustena märgitud funktsionaalsuse. Sellest räägiti lähemalt magistritöö viimases peatükis.

Loodud magistritöö on valdkondadeülene. Autori arvates oli huvitav endale selgeks teha bioloogiaalaseid tõdesid. Samuti näha antud valdkonna probleeme ja luua infotehnoloogilisi vahendeid nende lahendamiseks. Õpetlike kohti oli teisigi: veebiraamistike kasutamine, veebiteenuse loomine, hierarhiliste andmete hoiustamine relatsioonilises andmebaasis, kiirustestide läbi viimine ja nende põhjal meetodite optimeerimine, testide kirjutamine, nõuete kogumine jms.

Töös valminud taksonoomiamoodul on edukalt integreeritud PlutoF veebitöölauale ning

leiab kasutust ka teiste süsteemide taksonoomiamoodulina. Olulisemaks võiks pidada mooduli kasutuselevõttu üleeuroopalise projekti DINA raames loodavas loodusteaduslike kogude haldamise infosüsteemis.

Kasutatud kirjandus

- [1] V. Chavan ja P. Ingwersen, *Towards a data publishing framework for primary biodiversity data: challenges and potentials for the biodiversity informatics community*, *BMC Bioinformatics*, kd 10:S2, 2009.
- [2] C. Parr et al., *The encyclopedia of life v2: Providing global access to knowledge about life on earth*, *Biodiversity Data Journal*, kd 2:e1079, 2014.
- [3] E. Mayr ja W. J. Bock, *Journal of Zoological Systematics and Evolutionary Research*, kd 40:4, lk 169–194, 2002.
- [4] *Taxonomy*, 15-Oktoober-2008. [Internet]. <http://science.howstuffworks.com/life/biology-fields/taxonomy-info.htm>. (Viimati vaadatud: 22-Aprill-2015).
- [5] P. Halasz, *Biological classification*. [Internet]. http://commons.wikimedia.org/wiki/File:Biological_classification_L_Pengo.svg#file. (Viimati vaadatud: 22-Aprill-2015).
- [6] Vikipeedia, *Nomenklatuurikoodeks — vikipeedia*, 2013. [Internet]. <http://et.wikipedia.org/w/index.php?title=Nomenklatuurikoodeks&oldid=3557673>. (Viimati vaadatud: 22-Aprill-2015).
- [7] R. H. Whittaker, *Five-kingdom classification of life*. [Internet]. <http://www.britannica.com/media/full/107934>. (Viimati vaadatud: 22-Aprill-2015).
- [8] Y. Roskov et al., *Catalogue of life: 2014 annual checklist*. Leiden: Species 2000 & ITIS Catalogue of Life, 2014.
- [9] *The dDBJ/EMBL/GenBank feature table definition*, November-2014. [Internet]. <http://www.insdc.org/documents/feature-table>. (Viimati vaadatud: 22-Aprill-2015).
- [10] *International committee on nomenclature*. [Internet]. <http://www.bionomenclature.net/index.html>. (Viimati vaadatud: 22-Aprill-2015).
- [11] *Taxonomy*, 26-August-2014. [Internet]. <http://www.britannica.com/EBchecked/topic/584695/taxonomy/48695/The-objectives-of-biological-classification>. (Viimati vaadatud: 22-Aprill-2015).
- [12] J. David et al., *Biological nomenclature terms for facilitating communication in the naming of organisms*, *ZooKeys*, kd 192, lk 67–72, Mai 2012.

- [13] Wikipedia, *Taxonomic database — wikipedia, the free encyclopedia*, 11-December-2012. [Internet]. http://en.wikipedia.org/w/index.php?title=Taxonomic_database&oldid=527553758. (Viimati vaadatud: 22-Aprill-2015).
- [14] *Catalogue of life*, 15-Veebruar-2014. [Internet]. <http://www.catalogueoflife.org/col/info/about>. (Viimati vaadatud: 22-Aprill-2015).
- [15] D. Patterson, *About*, 03-September-2010. [Internet]. http://www.globalnames.org/about_GNA. (Viimati vaadatud: 22-Aprill-2015).
- [16] *EDIT platform for cybertaxonomy*. [Internet]. <http://cybertaxonomy.eu/>. (Viimati vaadatud: 22-Aprill-2015).
- [17] V. S. Smith et al., *Scratchpads: Concept*. [Internet]. <http://scratchpads.eu/about/concept>. (Viimati vaadatud: 22-Aprill-2015).
- [18] Wikipedia, *PostgreSQL — wikipedia, the free encyclopedia*, 2015. [Internet]. <http://en.wikipedia.org/w/index.php?title=PostgreSQL&oldid=653634995>. (Viimati vaadatud: 22-Aprill-2015).
- [19] Wikipedia, *Iterative design — wikipedia, the free encyclopedia*, 2014. [Internet]. http://en.wikipedia.org/w/index.php?title=Iterative_design&oldid=615095231. (Viimati vaadatud: 22-Aprill-2015).
- [20] B. Karwin, *SQL antipatterns - avoiding the pitfalls of database programming*. Raleigh, North Carolina: The Pragmatic Bookshelf, 2010, lk 34–53.
- [21] T. Tammet, *Veebiteenused reeglite ikkes, Arvutimaailm*, Veebruar 2007.
- [22] *Mis asi on veebiteenus?*, Detsember-2007. [Internet]. http://services.krediidiinfo.ee/wiki/index.php?title=Mis_asi_on_veebiteenus%3F&oldid=144. (Viimati vaadatud: 22-Aprill-2015).
- [23] D. Greenfield ja A. Roy, *Django packages: REST*, 2015. [Internet]. <https://www.djangopackages.com/grids/g/rest/>. (Viimati vaadatud: 22-Aprill-2015).
- [24] M. Lavin ja J. Elman, *Lightweight django*. Sebastopol, California: O'Reilly Media, 2014, lk 61.
- [25] G. Jansen, *Thoughts on RESTful API design*, 15-November-2012. [Internet]. <https://restful-api-design.readthedocs.org/en/latest/urls.html#url-structure>. (Viimati vaadatud:

22-Aprill-2015).

[26] M. Domenjoud, *Graph databases: an overview*, 2012. [Internet]. <http://blog.octo.com/en/graph-databases-an-overview/>. (Viimati vaadatud: 22-Aprill-2015).

Lisad

I. Terminid

Takson - organismirühm, millele on antud taksonoomiline nimi

Taksonoomiline nimi - ehk taksoni teaduslik nimi on kahest osast koosnev ladinakeelne nimi

Taksonoom - bioloog, kes töötab bioloogilise taksonoomia arendamisega

Taksonoomiline kategooria - tase taksonoomilises hierarhias, mida kasutatakse bioloogilises klassifikatsioonis. Taksonite kirjeldamise detailsus suureneb liikudes kõrgemalt kategoorialt madalamale.

Tavanimetus - ehk rahvapärane nimetus on nimi igapäevaselt kasutatavas keeles

Hübriid - ehk ristand on geneetiliselt erinevate vanemorganismide järglane

Nomenklatuur - nimetamise süsteem koos nimetamise reeglitega

Taxon - (plural taxa) a group of related organisms to which a taxonomic name is given.

Taxonomic name - (also scientific name) binomial latin name

Taxonomist - biologists who work practice the biological taxonomy - a sub-discipline of biology

Taxonomic rank - the level (relative position) in a taxonomic hierarchy used in biological classifications. Taxon's description on higher ranks is more general than on lower ranks.

Common name - (also vernacular name) is a name that is based on the normal language of everyday life

Hybrid - a mix of two animals, or plants of different breeds, varieties or genera

Nomenclature - a system of names or with rules of forming those names

Nomenklatuuri staatus - nime staatus, mis kirjeldab nime õigsust

Nomenclatural status - status of a name in nomenclature showing the validity of the name

LSID (Life Science Identifier) - loodusteaduse distsipliinides (elurikkuse informaatika, bioinformaatika) kasutatavad objektide unikaalsed identifikaatorid

LSID - unique life science identifiers used in biodiversity informatics and bioinformatics

II. Veebiteenuse sihtpunktid

URL	Kirjeldus	Meetod
/api/taxonomy/	veebiteenuse skeem	GET
/api/taxonomy/taxon/	Kõik võimalikud taksonid	GET
/api/taxonomy/taxon/	Uue taksoni loomine	POST
/api/taxonomy/taxon/<taksoni id>/	Antud taksoni uuendamine	PUT
/api/taxonomy/taxon/<taksoni id>/	Antud taksoni kustutamine	DELETE
/api/taxonomy/taxon/search/?search_query=<märksõna>	Taksonite otsimine otsimine teadusliku nime järgi	GET
/api/taxonomy/taxon/<taksoni id>/subtree/	Antud taksoni kõik lapsed	GET
/api/taxonomy/taxon/<taksoni id>/direct_children/	Antud taksoni kõik otsesed lapsed	GET
/api/taxonomy/taxon/<taksoni id>/higher_taxa/	Antud taksoni kõik vanemad	GET
/api/taxonomy/taxon/<taksoni id>/higher_taxa_intersection/?given_ids=<taksoni id>, <taksoni id>,...	Antud taksoni ja parameetritega kaasa antud taksonite ühised vanemad	GET
/api/taxonomy/vernacular_name/	Kõik olemasolevad tavanimetused	GET
/api/taxonomy/vernacular_name/	Uue tavanimetuse loomine	POST
/api/taxonomy/vernacular_name/<tavanimetuse id>/	Antud tavanimetuse uuendamine	PUT

URL	Kirjeldus	Meetod
/api/taxonomy/vernacular_name/ <tavanimetuse id>/	Antud tavanimetuse kustutamine	DELETE
/api/taxonomy/vernacular_name/ search/?search_query= <märksõ- na>	Tavanimetuse otsimine, mis tagastab ka viite taksonile	GET
/api/taxonomy/act/	Kõik olemasolevad aktid	GET
/api/taxonomy/act/<akti id>/	Antud akti andmed	GET
/api/taxonomy/taxon_rank/	Kõik olemasolevad kategooriad	GET
/api/taxonomy/taxon_rank/ <kate- gooria id>/	Antud kategooria andmed	GET
/api/taxonomy/language/	Kõik olemasolevad keeled	GET
/api/taxonomy/language/	Uue keele loomine	POST
/api/taxonomy/language/<keele id>/	Antud keele andmed	GET
/api/taxonomy/language/<keele id>/	Antud keele uuendamine	PUT
/api/taxonomy/language/<keele id>/	Antud keele kustutamine	DELETE
/api/taxonomy/hybridtaxonnode/	Kõik olemasolevad hübriidid (viide taksonile ja mõlemale vanemale)	GET
/api/taxonomy/hybridtaxonnode/	Uue hübriidi loomine	POST

URL	Kirjeldus	Meetod
/api/taxonomy/hybridtaxonnode/ <hübriidi id>/	Antud hübriidi andmed	GET
/api/taxonomy/hybridtaxonnode/ <hübriidi id>/	Antud hübriidi uuendamine	PUT
/api/taxonomy/hybridtaxonnode/ <hübriidi id>/	Antud hübriidi kustutamine	DELETE
/api/taxonomy/filter/	Kõik olemasolevad filtrid	GET
/api/taxonomy/filter/	Uue filtri loomine	POST
/api/taxonomy/filter/<filtri id>/	Antud filtri andmed	GET
/api/taxonomy/filter/<filtri id>/	Antud filtri uuendamine	PUT
/api/taxonomy/tree/	Kõik olemasolevad klassifikatsioonid	GET
/api/taxonomy/tree/	Uue klassifikatsiooni loomine	POST
/api/taxonomy/tree/ <klassifikat- siooni id>/	Antud klassifikatsiooni andmed	GET
/api/taxonomy/tree/ <klassifikat- siooni id>/	Antud klassifikatsiooni uuendamine	PUT
/api/taxonomy/tree/ <klassifikat- siooni id>/	Antud klassifikatsiooni kustutamine	DELETE
/api/taxonomy/tree/ <klassifikat- siooni id>/root_children/	Antud klassifikatsiooni kõige kõrge- mad taksonid	GET
/api/taxonomy/taxon_name __con- cept/	Kõik olemasolevad kontseptid	GET
/api/taxonomy/taxon_name __con- cept/<kontsepti id>/	Antud kontsepti andmed	GET

III. Veebiteenuse sihtpunktide väljade kirjeldused

api/taxonomy/taxon/

id - identifikaator

url - hüperlink objekti detailsele vaatele

parent - hüperlink objekti otsesele vanemale (null, kui tegu on sünonüümi või klassifikatsiooni esimese taseme taksoniga)

depth - kirjeldab taksonite arvu klassifikatsiooni tipu ja taksoni vahel

traverse_previous - hüperlink eelmisele taksonile klassifikatsioonis (kasutades eesjärjestust)

traverse_next - hüperlink järgmisele taksonile klassifikatsioonis (kasutades eesjärjestust)

synonyms - järjestus taksonitest, kellele takson on kehtiv nimi

parent_described_in - hüperlink vanemale, kus sünonüüm oli eelnevalt kirjeldatud

valid_name - hüperlink kehtivale taksonile

synonym_type - sünonüümi tüüp

nomenclatural_status - nomenklatuuristaatus

epithet - taksoni nimi või liigi nimi alates kindlast taksonoomilisest kategooriast

is_valid - defineerib, kas tegu on kehtiva taksoninimega

related_taxa - järjestus kõigist teistes klassifikatsioonides asetsevatest seotud taksonitest

hybrid_parents - järjestus taksonitest, mis moodustavad hübriid

taxon_name_concept - hüperlink seotud kontseptile

taxon_rank - hüperlink taksoni kategooriale

epithet_author - taksoni nime autor

year_described_in - aasta, millal takson kirjeldati

use_parentheses - märgib, kas ümarsulgusid tuleb kasutada taksoni täis nime kuvamisel

uid - LSID (seente puhul kasutatakse Index Fungorum'i poolt kirjeldatud LSID'sid)

code - kasutaja määratud kood

taxon_name - taksoni nimi

full_taxon_name - taksoni terviklik nimi, mis sisaldab autorit ja aastat (eriformaat, mis sõltub riigist)

tree - hüperlink klassifikatsioonile

acts - taksoniga seotud aktid

vernacular_names - järjestus taksoniga seotud tavanimetustest

citation_text - ainult sisestatav väli aktide jaoks (taksoni kuvamisel ei näidata)

remarks - ainult sisestatav väli aktide jaoks (taksoni kuvamisel ei näidata)

is_fossil - määrab, kas tegu on fossiiliga

api/taxonomy/vernacular_name/

id - identifikaator

url - hüperlink objekti detailsele vaatele

iso_639 - hüperlink keele objektile

common_name - tavanimetust
taxon_node - hüperlink seotud taksonile
is_preferred - määrab, kas tavanimetust peaks eelistama, kui taksonil on mitu tavanimetust

api/taxonomy/act/

id - identifikaator
url - hüperlink objekti detailsele vaatele
taxon_node - hüperlink taksoni detailsele vaatele
type - akti tüüp
previous_value - muutmisele eelnenud väärtus
remarks - märkmed, mis kirjeldavad taksoni lisamist või uuendamist
citation_text - kirjandusviide, mille põhjal takson lisati või seda uuendati
created_at - akti loomise aeg

api/taxonomy/taxon_rank/

id - identifikaator
abbreviation - kategooria lühend
zoology_rank - kategooria vastavalt rahvusvahelisele zooloogilisele nomenklatuurile (ICZN, <http://iczn.org/code>)
botany_rank - kategooria vastavalt rahvusvahelisele vetikate, seente, ja taimede nomenklatuurile (Melbourne Code, <http://www.iapt-taxon.org/nomen/>)

api/taxonomy/language/

iso_639 - ISO 639 standardile vastav kolmetäheline keele kood
url - hüperlink objekti detailsele vaatele
name - keele nimi

api/taxonomy/hybridtaxonnode/

id - identifikaator
url - hüperlink objekti detailsele vaatele
taxon_node - hüperlink hübriidi detailsele vaatele
hybrid_parent1 - hüperlink esimesele taksoni vanemale
hybrid_parent2 - hüperlink teisele taksoni vanemale

api/taxonomy/filter/

id - identifikaator

locale - hüperlink keele detailsele vaatele

lowest_rank - hüperlink taksonoomilise kategooria detailsele vaatele, millest madalmaid kategooriaid ei näidata

included_ranks - järjestus kategooriate hüperlinkidest, mida filtreerimisel sisse arvestada

excluded_ranks - järjestus kategooriate hüperlinkidest, mida filtreerimisel välja jätta

tree - hüperlink klassifikatsioonile, millele filter mõjub

api/taxonomy/tree/

id - identifikaator

url - hüperlink objekti detailsele vaatele

name - klassifikatsiooni nimi

origin_tree - hüperlink kloonimise aluseks olnud klassifikatsioonile

root_node - hüperlink kloonimisel aluseks olnud taksonile

api/taxonomy/taxon_name_concept/

id - identifikaator

url - hüperlink objekti detailsele vaatele

taxa - järjestus kõigist seotud taksonitest

IV. Kasutuskiiruse tulemused

Klassifi- katsiooni suurus	Päringu nimi	Iga mõõtmiskorra ajakulu					Keskmine aeg
1000	Kogu klassifikatsiooni pärimine (10 esimest vastet)	0.241	0.239	0.243	0.253	0.258	0.247
1000	Kogu klassifikatsioon pärimine (lehtedeks jaotamata)	1.684	1.759	1.710	1.692	1.750	1.719
1000	Klassifikatsiooni alamosa pärimine (10 esimest vastet)	0.228	0.232	0.231	0.236	0.234	0.232
1000	Otsingu teostamine (10 esimest vastet)	0.093	0.119	0.113	0.091	0.103	0.104
1000	Kõigi vanemate pärimine (10 esimest vastet)	0.068	0.072	0.070	0.070	0.074	0.071
1000	Kõigi vanemate ühisosa (10 esimest vastet)	0.074	0.069	0.069	0.076	0.072	0.072
1000	Otseste laste pärimine (10 esimest vastet)	0.143	0.153	0.141	0.148	0.152	0.147
1000	Taksonite filtreerimine keele põhjal (10 esimest vastet)	0.191	0.187	0.185	0.187	0.190	0.188
1000	Taksonite filtreerimine madalama kategooria põhjal (10 esimest vastet)	0.180	0.184	0.180	0.184	0.177	0.181
1000	Taksonite filtreerimine sisalduvate kategooriate põhjal (10 esimest)	0.187	0.182	0.184	0.186	0.176	0.183
1000	Taksonite filtreerimine välja jäetud kategooriate põhjal (10 esimest)	0.171	0.176	0.171	0.167	0.173	0.172
1000	Taksonite filtreerimine klassifikatsiooni põhjal (10 esimest vastet)	0.175	0.177	0.183	0.173	0.176	0.177
1000	Klassifikatsiooni alamosa ümber paigutamine (ühe taksoni keskmine aeg)	0.049	0.054	0.048	0.049	0.050	0.050
1000	Klassifikatsiooni kloonimine (ühe taksoni keskmine aeg)	0.014	0.016	0.015	0.016	0.015	0.015
10000	Kogu klassifikatsiooni pärimine (10 esimest vastet)	0.247	0.240	0.237	0.244	0.242	0.242
10000	Kogu klassifikatsioon pärimine (lehtedeks jaotamata)	1.774	1.766	1.794	1.793	1.734	1.772
10000	Klassifikatsiooni alamosa pärimine (10 esimest vastet)	0.232	0.237	0.232	0.236	0.230	0.233
10000	Otsingu teostamine (10 esimest vastet)	0.096	0.088	0.106	0.127	0.096	0.103
10000	Kõigi vanemate pärimine (10 esimest vastet)	0.079	0.087	0.082	0.085	0.086	0.084
10000	Kõigi vanemate ühisosa (10 esimest vastet)	0.086	0.091	0.095	0.089	0.089	0.090
10000	Otseste laste pärimine (10 esimest vastet)	0.142	0.144	0.153	0.146	0.143	0.146
10000	Taksonite filtreerimine keele põhjal (10 esimest vastet)	0.192	0.191	0.197	0.186	0.196	0.192
10000	Taksonite filtreerimine madalama kategooria põhjal (10 esimest vastet)	0.185	0.183	0.181	0.185	0.186	0.184
10000	Taksonite filtreerimine sisalduvate kategooriate põhjal (10 esimest)	0.184	0.185	0.178	0.179	0.184	0.182
10000	Taksonite filtreerimine välja jäetud kategooriate põhjal (10 esimest)	0.189	0.188	0.181	0.194	0.191	0.189
10000	Taksonite filtreerimine klassifikatsiooni põhjal (10 esimest vastet)	0.184	0.178	0.182	0.178	0.186	0.182
10000	Klassifikatsiooni alamosa ümber paigutamine (ühe taksoni keskmine aeg)	0.050	0.048	0.051	0.049	0.053	0.050
10000	Klassifikatsiooni kloonimine (ühe taksoni keskmine aeg)	0.014	0.015	0.015	0.016	0.015	0.015
100000	Kogu klassifikatsiooni pärimine (10 esimest vastet)	0.267	0.275	0.270	0.273	0.357	0.288
100000	Kogu klassifikatsioon pärimine (lehtedeks jaotamata)	1.798	1.710	1.745	1.730	1.767	1.750
100000	Klassifikatsiooni alamosa pärimine (10 esimest vastet)	0.238	0.235	0.233	0.230	0.232	0.234
100000	Otsingu teostamine (10 esimest vastet)	0.132	0.095	0.099	0.103	0.113	0.108
100000	Kõigi vanemate pärimine (10 esimest vastet)	0.087	0.083	0.084	0.083	0.085	0.084
100000	Kõigi vanemate ühisosa (10 esimest vastet)	0.086	0.092	0.087	0.085	0.086	0.087
100000	Otseste laste pärimine (10 esimest vastet)	0.148	0.146	0.145	0.148	0.147	0.147
100000	Taksonite filtreerimine keele põhjal (10 esimest vastet)	0.225	0.208	0.210	0.206	0.215	0.213
100000	Taksonite filtreerimine madalama kategooria põhjal (10 esimest vastet)	0.177	0.178	0.185	0.184	0.178	0.180
100000	Taksonite filtreerimine sisalduvate kategooriate põhjal (10 esimest)	0.180	0.189	0.179	0.182	0.187	0.183
100000	Taksonite filtreerimine välja jäetud kategooriate põhjal (10 esimest)	0.218	0.215	0.202	0.213	0.227	0.215
100000	Taksonite filtreerimine klassifikatsiooni põhjal (10 esimest vastet)	0.200	0.200	0.198	0.204	0.197	0.200
100000	Klassifikatsiooni alamosa ümber paigutamine (ühe taksoni keskmine aeg)	0.050	0.051	0.049	0.049	0.049	0.050
100000	Klassifikatsiooni kloonimine (ühe taksoni keskmine aeg)	0.015	0.015	0.015	0.015	0.015	0.015

V. Litsents

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Siim Halapuu (sünnikuupäev: 26. august 1989)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose Eluslooduse klassifikatsioonide haldamise andmebaas ja veebiliides, mille juhendajad on Kessy Abarenkov ja Vambola Leping,
 - 1.1 reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
 - 1.2 üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.
3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, 2. mai 2015. a.